Regression Lines

Use *y* on *x* when:

- *x* is controlled (not random) variable
- *or* both *x* and *y* are random and you are estimating *y* from given *x*

3 In an agricultural experiment, the relationship between the amount of water supplied, x units, and the yield, y units, was investigated. Six values of x were chosen and for each value of x the corresponding value of y was measured. The results are shown in the table.

x	1	2	3	4	5	6
y	3	6	8	8	11	10

(i) Give a reason why the regression line of x on y is not suitable in this context. **[1]**

Binomial & Geometric Distributions – Modelling Assumptions

- For each trial, the probability of a success is constant
- For each trial, the events "success" and "failure" are independent of the corresponding events for every other trial.

These must be contextualised!

They are not the same!

Not a statement about parameter values

1 Andy makes repeated attempts to thread a needle. The number of attempts up to and including his first success is denoted by X.

(i) State two conditions necessary for X to have a geometric distribution. **[2]**

(iii) Suggest a reason why one of the conditions you have given in part **(i)** might not be satisfied in this context. **[2]**

7 At a factory that makes crockery the quality control department has found that 10% of plates have minor faults. These are classed as 'seconds'. Plates are stored in batches of 12. The number of seconds in a batch is denoted by X.

(i) State an appropriate distribution with which to model X. Give the value(s) of any parameter(s) and state any assumptions required for the model to be valid. **[4]**

Poisson Distribution – Modelling Assumptions

- Events occur at constant average rate
- Events occur independently of one another.

These must be contextualised!

"Singly" is a subset of "independently".

Not a statement about parameter values.

5 In a large region of derelict land, bricks are found scattered in the earth.

(i) State two conditions needed for the number of bricks per cubic metre to be modelled by a Poisson distribution. **[2]**

Random Sampling

- Every element of the population is equally likely to be chosen
- Each element is chosen independently

People's opinions etc don't have to be independent – it's the selection that matters.
These conditions match those for a binomial distribution.

4 A survey is to be carried out to draw conclusions about the proportion p of residents of a town who support the building of a new supermarket. It is proposed to carry out the survey by interviewing a large number of people in the high street of the town, which attracts a large number of tourists.

(i) Give two different reasons why this proposed method is inappropriate. **[2]**

(ii) Suggest a good method of carrying out the survey. **[3]**

(iii) State two statistical properties of your survey method that would enable reliable conclusions about p to be drawn. **[2]**

Central Limit Theorem

- Regardless of the shape of the parent distribution, the mean of a sufficiently large sample is approximately normally distributed

Not a statement about parameter values
Distinguish between "Necessary to use the CLT" and "Possible to use the CLT".

6 The continuous random variable R has the distribution $N(\mu, \sigma^2)$. The results of 100 observations of R are summarised by

$$\Sigma r = 3360.0, \quad \Sigma r^2 = 115\,782.84.$$

(i) Calculate an unbiased estimate of μ and an unbiased estimate of σ^2. **[4]**

(ii) The mean of 9 observations of R is denoted by $\overline{R}$. Calculate an estimate of $P(\overline{R} > 32.0)$. **[4]**

(iii) Explain whether you need to use the Central Limit Theorem in your answer to part **(ii)**. **[2]**

Continuity Corrections

- When approximating to a discrete distribution using a continuous

Common: normal approximation to binomial or Poisson
Less common: mean of sample chosen from discrete distribution such as uniform
Here the continuity correction is $\pm 1/2n$, but it's much easier to look at the distribution of the *sum* rather than the *mean*.

7 The continuous random variable T is equally likely to take any value from 5.0 to 11.0 inclusive.

(i) Sketch the graph of the probability density function of T. **[2]**

(ii) Write down the value of $E(T)$ and find by integration the value of $Var(T)$. **[5]**

(iii) A random sample of 48 observations of T is obtained. Find the approximate probability that the mean of the sample is greater than 8.3, and explain why the answer is an approximation. **[6]**